GRAPHICAL INTERFACE OF THE *SCIMAGO JOURNAL AND COUNTRY RANK*: AN INTERACTIVE APPROACH TO ACCESSING BIBLIOMETRIC INFORMATION

Yusef Hassan-Montero, Vicente P. Guerrero-Bote y Félix De-Moya-Anegón

Nota: Este artículo puede leerse traducido al español en:
http://www.elprofesionaldelainformacion.com/contenidos/2014/may/07_esp.pdf

Yusef Hassan-Montero has a PhD in Documentation from the *University of Granada* and works for *SCImago Lab* in the design and development of interfaces for scientific visualization. He directs *nosolousabilidad.com* online magazine, is co-organizer of *UXSpain*, the most important national event on user experience in Spain, and teaches about interaction design in graduate courses of the *Open University of Catalonia* (*UOC*) and *Pompeu Fabra University*.
http://orcid.org/0000-0002-4314-1175

Churriana de la Vega, 18194 Granada, Spain
yusefhassan@gmail.com

Vicente P. Guerrero-Bote received a Ph.D. in Information Sciences in the *University of Granada*; a B.Sc. in Physics in the *Complutense University of Madrid*; is university professor of Research Evaluation and Network Analysis at the *University of Extremadura*, where he is the head of the *Department of Information and Communication*. He is a member of *SCImago, Research Group* dedicated to Scientometrics, being one of the developers of the *SCImago Journal Rank*.
http://orcid.org/0000-0003-4821-9768

Facultad de Ciencias de la Documentación y la Comunicación, *Universidad de Extremadura*
Pl. Ibn Marwan, s/n. 06001 Badajoz, Spain
guerrero@unex.es

Félix De-Moya-Anegón, Ph.D. in History from the *University of Granada*, is a research professor in the *Institute of Public Goods and Policies* of the *Higher Council for Scientific Research* (*CSIC*) in Madrid, and principal researcher of the associated unit *SCImago Research Group*. He has directed the R&D projects *SCImago Journal & Country Rank* (*SJR*), *SCImago Institutions Rankings* (*SIR*), *Atlas of Science*, etc.
http://orcid.org/0000-0002-0255-8628

Centro de Ciencias Humanas y Sociales (*CCHS*), *CSIC*
Albasanz, 26-28, 28037 Madrid, Spain
SCImago Research Group
felix.moya@scimago.es

Abstract

A graphical interface designed to access the bibliometric indicators database of the *SCImago Journal & Country Rank* portal (*http://scimagojr.com/shapeofscience*) is described. The map was generated from the relational matrix based on the citation, co-citation, and bibliographic coupling formed by the nearly 20,000 publications (journals and conference proceedings) registered in *Scopus*. The map layout is based on a variant of a force-directed algorithm using Noack's proposal of an edge-repulsion energy model. The interface allows the publications' bibliometric indicators and the cluster structures that they form to be displayed based on their shared use by the authors of the documents. To facilitate navigation, the interface automatically positions the reference areas and subject categories which are viewable via zoom-and-pan. The interface may be found to constitute a useful tool for analyses of the *Scopus* publications' presence in different scientific domains, and of the global distribution of the publishing capacity of different countries and regions. It uses the method of overlaying maps to locate subsets of selected publications in the context of the global publication structure.

Keywords

Maps of scientific publications, Graphical interfaces, Bibliometric information, Science maps, Science analysis tools, Citations, *Scopus*, *SCImago*.

Título: Interfaz gráfica del *SCImago Journal & Country Rank:* Un enfoque interactivo para acceder a la información bibliométrica

Article received on: 20-05-2014
Acepted on: 28-05-2014

Resumen

Diseño de una interfaz visual para el acceso a la base de datos de indicadores bibliométricos del portal *SCImago Journal & Country Rank*. El mapa se genera a partir de la matriz relacional basada en citación, cocitación y *bibliographic coupling* que forman las casi 20.000 publicaciones (revistas y actas de congreso) registradas en *Scopus*. El *layout* del mapa se basa en una variante de un algoritmo *force-directed* mediante el modelo de energía *edge-repulsion* propuesto por Noack. La interfaz permite la visualización de los indicadores bibliométricos de las publicaciones y la estructura de clusters que forman en función de su uso compartido por los autores de los trabajos. Para facilitar la navegación la interfaz incluye la referencia de áreas y categorías temáticas mediante posicionamiento automático y visibles mediante *zoom+pan*. La interfaz puede ser una herramienta útil para analizar la presencia de las publicaciones *Scopus* en los diferentes dominios científicos, así como la distribución mundial de la capacidad editorial de los diferentes países o regiones. La interfaz utiliza la metodología de mapas *overlay* para situar los subconjuntos de publicaciones seleccionadas en el contexto de la estructura global de publicaciones.

Palabras clave

Mapas de las publicaciones científicas, Mapas de la ciencia, Datos bibliométricos, Citas, Interfaces gráficas, Representación de la ciencia, Herramientas para analizar la ciencia, *Scopus*, *SCImago*.

Hassan-Montero, Yusef; **Guerrero-Bote, Vicente P.**; **De-Moya-Anegón, Félix** (2014). "Graphical interface of the *SCImago Journal and Country Rank*: An interactive approach to accessing bibliometric information". *El profesional de la información*, May-June, v. 23, n. 3, pp. 272-278.

http://dx.doi.org/10.3145/epi.2014.may.07

Introduction

As noted by **Van-Eck** (2011), bibliometric maps representing the relationships among nodes can mainly take two forms: as graphical links (such as lines or arcs), or, without such links, by positioning the nodes to represent their mutual proximity or distance. For large networks with a high density of links, as in the case of the present study, the latter option, in which the degree of similarity or clustering among nodes is reflected by their spatial proximity, is the only viable option to ensure that the map is legible and understandable.

In the construction of distance-based maps, the algorithmic technique most widely used is known as MDS (multi-dimensional scaling). This technique takes as input a similarity or distance matrix in an N-dimensional space that is reduced to two or three dimensions (coordinates), trying to make the distances between nodes in the final representation approximate as closely as possible the distance they had in the N-dimensional space. The main drawback of MDS, however, is its tendency to produce circular maps where the nodes of higher degree are located near the centre of the map (**Van-Eck**; **Waltman**, **Dekker**; **Van-den-Berg**, 2010), an artefact of the logic of its algorithmic procedure that distorts the structural realities of the network.

An alternative to MDS which does not impose this artificial structure on the final layout, and in which the clusters of nodes are more easily recognized, is VOS (visualization of similarities) (**Van-Eck**; **Waltman**, 2007), used successfully in constructing various bibliometric maps. The two techniques are closely related. Indeed, VOS may be considered a weighted type of MDS in which proximities and weights are determined in one particular way (**Van Eck**; **Waltman**; **Dekker**; **Van-den-Berg**, 2010).

Another alternative is to use algorithms of a force-directed or energy-based type, which iteratively apply forces of attraction and repulsion among the nodes to reach a state (i.e., layout) of minimum energy. Many of the more popular force-directed algorithms (**Eades**, 1984; **Davidson**; **Harel**, 1989; **Fruchterman**; **Reingold**, 1991) are unsuitable for constructing distance-based bibliometric maps since their final layout mainly favours æsthetic criteria such as a uniform distribution of the nodes, the optimal use of space, or the avoidance of crossing between links. I.e., these algorithms do not seek to reflect the degree of similarity among nodes through their position, but to obtain readable layouts in representations based on graphical links. While the algorithm of **Kamada** & **Kawai** (1989) does reflect certain properties of the network in the layout, such as the geodetic distance between the nodes of the graph, it too imposes artificial æsthetic criteria such as uniform link length, and therefore equal distances between nodes.

> Force-directed methods have two components: an energy model that defines which layout to compute, and an energy minimization algorithm which defines how to compute it

In contrast, a proposal that is especially well-suited to the construction of distance-based bibliometric maps is that of **Noack**'s force-directed approach (2004, 2007). Force-directed methods have two components: an energy model that defines which layout to compute, and an energy minimization algorithm which defines how to compute it. **Noack**'s proposal focuses on the former, using two energy models termed node-repulsion LinLog and edge-repulsion LinLog. The main quality of these models, especially the edge-repulsion model, lies in their ability to yield layouts in which the position of the nodes not only reveals similarity relations, but also the community structure (**Noack**, 2009).

Figure 1. General view of the layout with the interface. The Subject Areas were selected as labels, and the *SJR-2012* values as the size of the nodes.

using the edge-repulsion energy model proposed by **Noack** (2007).

Given the size of the network in this study, in order to process its layout with reasonable efficiency, we used a multi-scale strategy (**Hadany**; **Harel**, 2001). Such strategies are based on computing an abstraction of the graph on which a general layout is calculated, to which details are then added gradually, editing and fine-tuning the layout. The procedure used in the present study consisted of two phases: coarsening and un-coarsening.

Regarding the quality of the final layout, **Van-Eck** (2001) states *"In the case of maps with lots of objects, the accuracy of the VOS mapping technique at the local level can be somewhat disappointing"*, and adds *"An interesting mapping technique that seems to yield accurate results both at the local and at the global level is the LinLog technique"*. Finally, as will be discussed below, the force-directed type of algorithms have the advantage of being easily tunable through the application of well-known optimization strategies.

The objective of the present study was to design a graphical interface for *SJ&CR* based on the relationships connecting scientific journals, so that use of an overlay method would permit the academic production of different geographic and thematic domains to be represented on the overall base map.

Methods and data

Bibliometric maps display the structure of a domain through a graphical representation of the units of analysis and of their relationships. In the present study, the units of analysis are international scientific journals. To calculate the degree to which these journals are related, we applied the mean of the normalized values of three measures: citation, co-citation, and bibliographic coupling. When there only existed bibliographic coupling between two journals, the link is discarded due to its relative insignificance.

To facilitate subsequent processing, the directed graph formed by the journals is transformed into an undirected graph in which the weight of each link between two nodes is the sum of the weights of their directed links. The resulting network is composed of 19,540 vertices and 3,612,061 links.

The network was generated from *Scopus* data (2003-2012) and from the thematic classification of journals used in *SJ&CR*. This latter consists of Subject Areas (26 plus the General class) and Specific Subject Areas or Categories (309).

We started by implementing the force-directed algorithm

The coarsening phase takes as input the original graph G_0. In this graph, a downwards traverse is made of all its links ordered by their density, which is defined as

$$d_{ij} = w_{ij} / wD_i \bullet wD_j$$

where w_{ij} is the weight of the link, and wD_i and wD_j are the weighted degrees of the nodes (sum of the weights of all their links). For each link, if neither of its nodes has previously been collapsed, they are then collapsed to form a single node in graph G_1, whose links will be the result of aggregating the links of the original two nodes. This process is repeated until reaching a graph G_n, the graph of greatest abstraction, in which it is no longer possible to continue collapsing nodes.

> The results obtained in the present study suggest that the use of multi-scale strategies in combination with LinLog energy models, is a very efficient method to achieve high quality layouts for large networks

The un-coarsening phase starts from graph G_n, to which the force-directed positioning algorithm is applied using a very small number of iterations (in fact, we only used two iterations, although this is a criterion that depends on the topology and size of the network). After applying the algorithm, the positions (coordinates) of each node are inherited by the two nodes it corresponded to in graph G_{n-1}. This process of positioning or refining the layout is repeated until the original graph G_0 is reached again.

Although **Noack** (2007) shows a certain reluctance to regard as appropriate the use of multi-scale strategies in combination with LinLog energy models, the results obtained in the

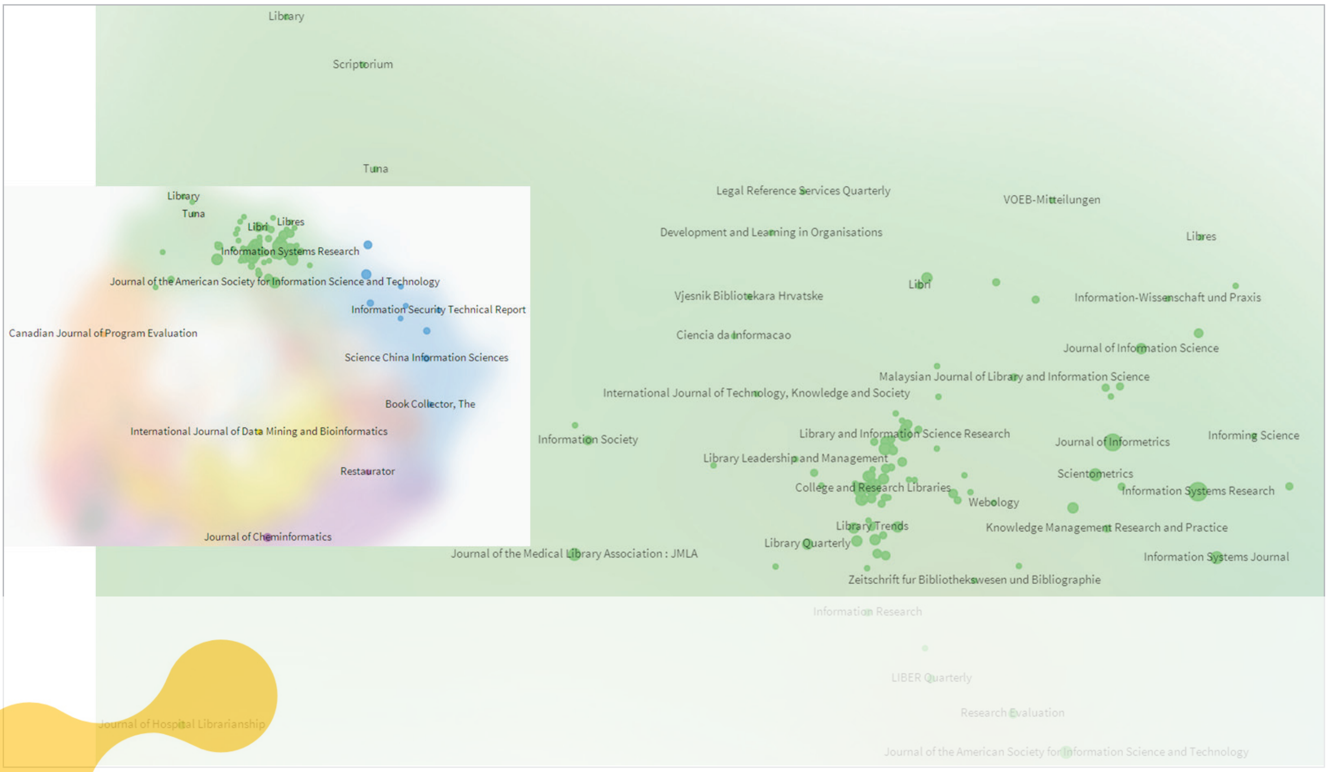

Figure 2. Library & Information Science worldwide with the *SJR-2012* as the absolute size of the nodes.

present study suggest that the procedure is indeed a very efficient method with which to achieve high quality layouts for large networks.

The nodes are represented graphically by discs whose size (their area) encodes the value of whichever bibliometric indicator that the user has selected (*SJR*, documents, references, citations...). The colour of the node identifies the cluster to which the journal belongs. For the calculation of the different clusters, we used the algorithm implemented by **Noack** (2007) based on **Newman** and **Girvan**'s measure of modularity (**Newman**, 2004; **Newman**; **Girvan**, 2004).

Once the overall map has been generated for the corresponding period, to select a domain, the interface provides the user with different options: the editor's country, the

Figure 3. Library & Information Science in Spain with *SJR-2012* as the absolute size of the nodes.

authors' country, the area, and the subject category. The results of the selected domain are superimposed using the overlay technique (**Leydesdorff**; **De-Moya-Anegón**; **Guerrero-Bote**, 2010; 2014). In this case, as well as generating the overall map with different procedures and data, with everything being done online, the clusters are not only coded by the colour of the nodes, but also by the contour formed by the nodes of each cluster, combining Gaussian blur with transparencies.

To enable detailed exploration of the map, an interactive zoom-and-pan mechanism is included, as well as a journal search on title.

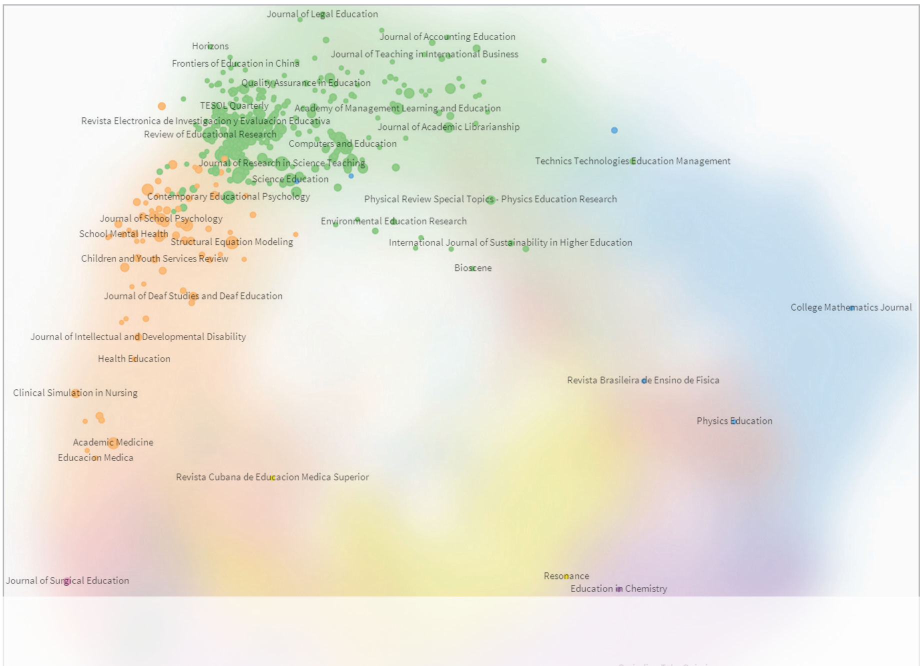

Figure 4. Overlay of the journals of the Category of Education with the *SJR-2012* as node size

To facilitate the interpretation of the map, the interface allows one to select showing either the journal titles or the area and subject category labels. Each of these labels is positioned at the centroid of all the journals covering that area or category. Overlap between labels is avoided using the mechanism described in **Koh**, **Lee**, **Kim** & **Seo** (2010).

Results and discussion

Figure 1 presents an overview of the interface with the general layout. It includes all the journals (19,540), represented by nodes of sizes proportional to *SJR-2012*. As can be seen in the panel on the left, one can choose the indicator that will be rendered with the size of the nodes. This may be general for the journal, as is the case with *SJR*, or it may depend on the domain that has been selected, as might be the output of the domain in the journal in question. This value can also be set to either absolute or relative.

The labels can also be set to either Subject Areas or Journal Titles.

The nine colours of the nodes correspond to the nine clusters or communities detected by the algorithm. The green cluster at the top is devoted mainly to Social Sciences and Humanities. The orange cluster on the left corresponds to Psychology, Neuroscience, Nursing, and Health Professions. Below it is a dark pink cluster devoted to Medicine which merges at the bottom with a blue-green cluster devoted to Ophthalmology and Optometry and then a teal-coloured cluster dedicated to Dentistry. More to the right, there are two clusters, one yellow dedicated to Life Sciences and the other purple devoted to Chemistry, Materials Science, and Chemical Engineering. Up and to the right, these two clusters fuse with another two, one reddish brown of Earth and Planetary Sciences, and the other blue, and rather elongated, starting from the bottom with Physics and Astronomy, then passing up through Mathematics, then Computer Science, to end with Decision Sciences when it merges with the first cluster.

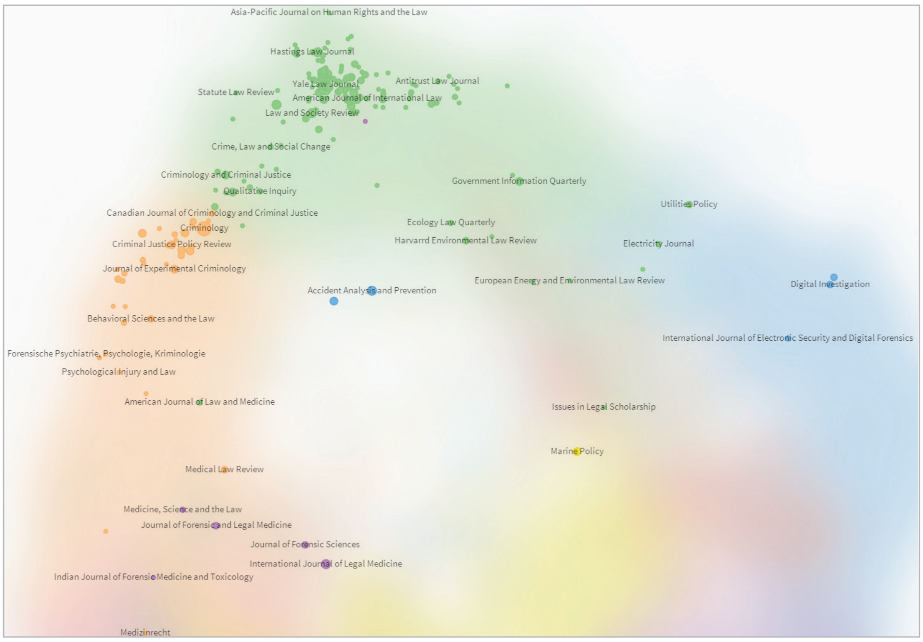

Figure 5. Overlay of the journals of the Category of Law with the *SJR-2012* as node size.

This circular layout is consistent with the general consensus of other maps of the sciences (**Klavans**; **Boyack**, 2009; **Leydesdorff**; **De-Moya-Anegón**; **Guerrero-Bote**, 2010; 2014). The goodness of

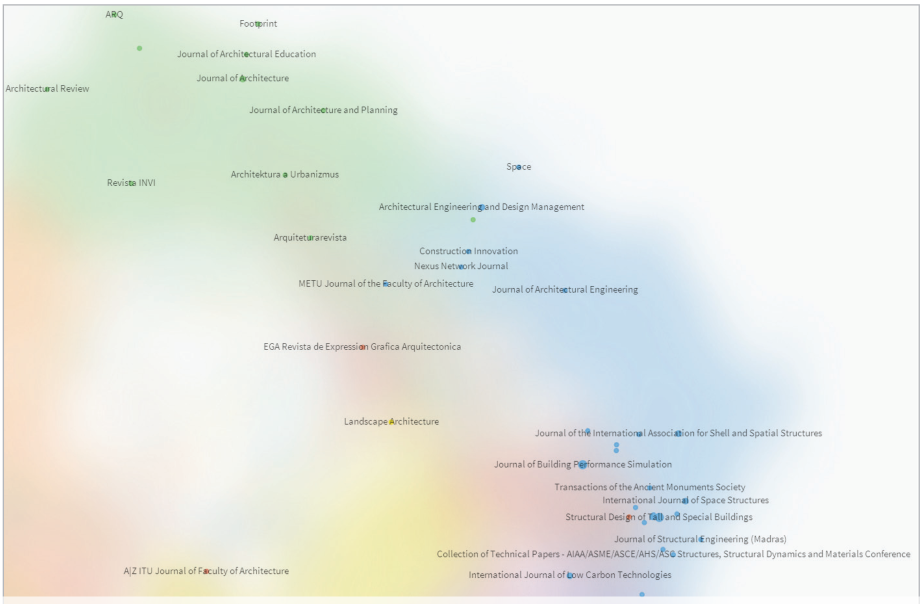

Figure 6. Overlay of the journals of the Category of Architecture with the *SJR-2012* as node size

the layout is evident in the very logic of the placement of the Subject Areas. Also worth mentioning is the clear coordination between the clustering algorithm and the layout since by far the greater part of the journal clusters (identified by their colours) are also grouped together visually.

If we zoom in on any of the parts with the Subject Areas as labels, those labels get larger, but they also blur out while the Specific Subject Area or Category labels start to appear.

Figure 2 shows the overlay corresponding to journals of the Specific Subject Area of Library & Information Science worldwide. The figure includes both a general view and a zoom-in on the most prominent grouping in which one can see a group of journals with Humanistic, Economic, and Scientometric content. Figure 3 shows the same Category but now delimited to the domain of Spain. There are now fewer journals since those displayed include only journals in which Spanish authors published in 2012. With fewer journals, some Spanish-language journals, such as *Scire* and the *Revista Española de Documentación Científica*, have become visible, and others can be located even though their title is not displayed because of the overlap, examples being the *Revista General de Información y Documentación* (slightly below and to the left of the *Revista Española de Documentación Científica*), *Investigación Bibliotecnológica* (a little further down), *El Profesional de la Información* (and a little more further down), or *Cybermetrics* (above the *Journal of Informetrics*).

Figure 4 shows the overlay corresponding to the journals of the Category of Education worldwide. One can see the main grouping in the border zone between Psychology and Social Sciences. There also appear some scattered journals that correspond to educational journals in various specific disciplines.

Something similar is the case in figure 5 for Law. One sees the main grouping of journals in Humanities. There is another group of journals on Criminology found within Psychology, and, at the bottom, some journals devoted to Forensic Science and others that have published papers devoted to the different branches of Law.

In Architecture (figure 6), one sees two groups of journals, one very technical, and close to Engineering and the Hard Sciences, and another dedicated to architectural art.

> "Today it is necessary to combine different techniques, and especially to seek applications of the representations generated that go beyond what is obvious analytically"

The case of Biomedical Engineering (figure 7) is somewhat special. Although *Scopus* includes it within Engineering, its journals do not focus on this area. Indeed, the label of the Category at the centroid of the positions of its journals appears at the bottom. Many journals are in a purple zone, corresponding to Chemistry or Materials Science; others in a yellow zone, corresponding to Biomedicine; and others in a pink zone, more within Medicine. This may be evidence

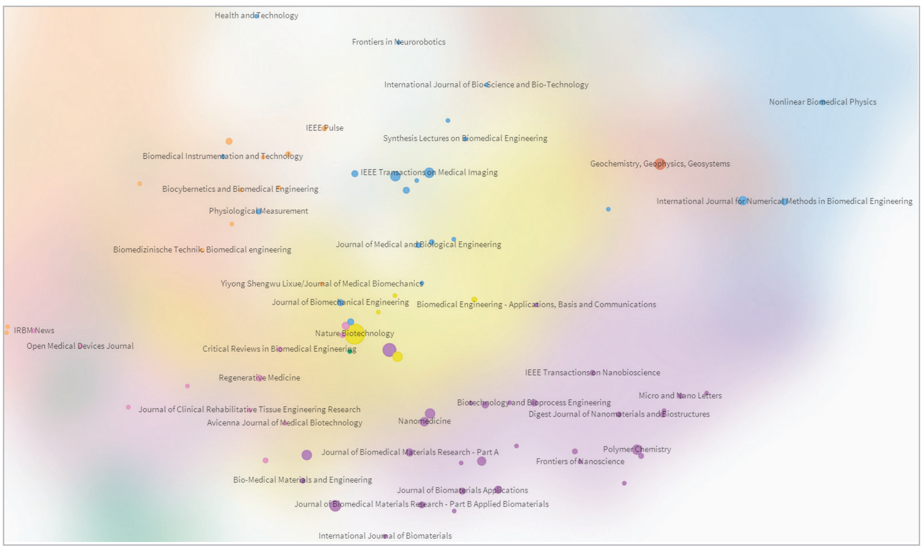

Figure 7. Overlay of the journals of the Category of Biomedical Engineering with the *SJR-2012* as node size

of a bad placement of the Category within the Engineering category in *Scopus*.

Conclusions

We have described a combination of force-directed, clustering, and overlay techniques to generate interactive maps capable of being used as efficient interfaces to access bibliometric information. Gone are the days when the mastery of a technique of mapping some particular form of bibliographic information was a novelty in the field of research. Today it is necessary to combine different techniques, and especially to seek applications of the representations generated that go beyond what is obvious analytically. Nonetheless, our principal objective with the present development was to provide users with a tool that would allow them to interact visually with all the information accessible through the *SCImago Journal & Country Rank* portal. The aim was, therefore, to offer an alternative to the traditional form-based access routes that are usual in bibliographic databases. This objective could not be attained without solving a host of human-computer interaction (HCI) problems posed by the simple volume of information being handled, and by the need to work at different levels of zoom with classifications represented graphically according to the user's requested context.

The solutions that we have described take as starting point the existence of a database of bibliometric indicators grouped by geographic domains, scientific fields, and publications which can now also be analysed visually in the form of graphical displays. This graphical approach is intended to constitute just another way of analysing that part of reality which the indicators can shed light on. In no way does it pretend to replace the tabular, georeferenced, or any other type of approach that the user can take. The methodological principle that motivates the need to generate multiple representations of the same social reality is that of their analytical complementarity.

Acknowledgments

This work was financed by the *Junta de Extremadura, Consejería de Empleo, Empresa e Innovación* and by the *European Social Fund* as part of the research group grant GR10019.

References

Davidson, Ron; **Harel, David** (1989). *Drawing graphs nicely using simulated annealing*. Technical Report CS89-13, Department of Applied Mathematics and Computer Science, The Weizmann Institute, Rehovot, Israel.

Eades, Peter (1984). "A heuristic for graph drawing". *Congressus numerantium*, v. 42, n. 11, pp. 149-160.

Fruchterman, Thomas M. J.; **Reingold, Edward M.** (1991). "Graph drawing by force-directed placement". *Software - practice & experience* (Wiley), v. 21, n. 11, pp. 1129-1164.
http://dx.doi.org/10.1002/spe.4380211102

Hadany, Ronny; **Harel, David** (2001). "A multi-scale algorithm for drawing graphs nicely". *Discrete applied mathematics*, v. 113, n. 1, pp. 3-21.
http://dx.doi.org/10.1016/S0166-218X(00)00389-9

Kamada, Tomihisa; **Kawai, Satoru** (1989). "An algorithm for drawing general undirected graphs". *Information processing letters*, n. 31, pp. 7-15.
http://dx.doi.org/10.1016/0020-0190(89)90102-6

Klavans, Richard; **Boyack, Kevin W.** (2009). "Towards a consensus map of science". *Journal of the American Society for Information Science and Technology*, v. 60, n. 3, pp. 455-476.
http://dx.doi.org/10.1002/asi.20991

Koh, Kyle; **Lee, Bongshin**; **Kim, Bohyoung**; **Seo, Jinwook** (2010). "ManiWordle: Providing flexible control over Wordle". *IEEE Trans vis comput graph.*, v. 16, n. 6, pp. 1190-1197.
http://dx.doi.org/10.1109/TVCG.2010.175

Leydesdorff, Loet; **De-Moya-Anegón, Félix**; **Guerrero-Bote, Vicente P.** (2010). "Journal maps on the basis of *Scopus* data: A comparison with the *Journal Citation Reports* of the *ISI*". *Journal of the American Society for Information Science and Technology*, v. 61, n. 2, pp. 352-369.
http://dx.doi.org/10.1002/asi.21250

Leydesdorff, Loet; **De-Moya-Anegón, Félix**; **Guerrero-Bote, Vicente P.** (2014). "Journal maps, interactive overlays, and the measurement of interdisciplinarity on the basis of *Scopus* data (1996-2012)". *Journal of the American Society for Information Science and Technology* (en prensa), Wiley.
http://www.arxiv.org/pdf/1310.4966

Newman, Mark E. J. (2004). "Analysis of weighted networks". *Physical review E*, n. 70, 056131.
http://dx.doi.org/10.1103/PhysRevE.70.056131

Newman, Mark E. J.; **Girvan, Michelle** (2004). "Finding and evaluating community structure in networks". *Phys. rev. E*, n. 69, 026113.
http://dx.doi.org/10.1103/PhysRevE.69.026113

Noack, Andreas (2004). *Visual clustering of graphs with nonuniform degrees*. Technical report 02/04, Brandenburg University of Technology at Cottbus (BTU).

Noack, Andreas (2007). "Energy models for graph clustering". *Journal of graph algorithms and applications*, v. 11, n. 2, pp. 453-480.
http://dx.doi.org/10.7155/jgaa.00154

Noack, Andreas (2009). "Modularity clustering is force-directed layout". *Physical review E*, n. 79, 026102.
http://dx.doi.org/10.1103/PhysRevE.79.026102

Van-Eck, Nees-Jan (2011). *Methodological advances in bibliometric mapping of science*. ERIM PhD Series research in management, 247. ISBN: 978 90 5892 291 5

Van-Eck, Nees-Jan; **Waltman, Ludo** (2007). "VOS: a new method for visualizing similarities between objects". In: H.-J. Lenz; R. Decker (eds.), *Advances in data analysis: Proceedings of the 30th annual conference of the German Classification Society* (pp. 299-306).

Van-Eck, Nees-Jan; **Waltman, Ludo**; **Dekker, Rommert**; **Van-den-Berg, Jan** (2010). "A comparison of two techniques for bibliometric mapping: Multidimensional scaling and VOS". *Journal of the American Society for Information Science and Technology*, v. 61, n. 12, pp. 2405-2416.
http://dx.doi.org/10.1002/asi.21421